

# DISCUSSION PAPER SERIES

No. 3527

## ENCLAVES, LANGUAGE AND THE LOCATION CHOICE OF MIGRANTS

Thomas Bauer, Gil S Epstein and Ira Gang

*LABOUR ECONOMICS*



**C**entre for **E**conomic **P**olicy **R**esearch

[www.cepr.org](http://www.cepr.org)

Available online at:

[www.cepr.org/pubs/dps/DP3527.asp](http://www.cepr.org/pubs/dps/DP3527.asp)

# ENCLAVES, LANGUAGE AND THE LOCATION CHOICE OF MIGRANTS

**Thomas Bauer**, IZA, Bonn and CEPR  
**Gil S Epstein**, Bar-Ilan University, Ramat-Gan and CEPR  
**Ira Gang**, Rutgers University

Discussion Paper No. 3527  
September 2002

Centre for Economic Policy Research  
90–98 Goswell Rd, London EC1V 7RR, UK  
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999  
Email: [cepr@cepr.org](mailto:cepr@cepr.org), Website: [www.cepr.org](http://www.cepr.org)

This Discussion Paper is issued under the auspices of the Centre's research programme in **LABOUR ECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Economic and Social Research Council, under which an ESRC Resource Centre operates within CEPR; the Esmée Fairbairn Charitable Trust; and the Bank of England. These organizations do not give prior review to the Centre's publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Thomas Bauer, Gil S Epstein and Ira Gang

September 2002

## ABSTRACT

### Enclaves, Language and the Location Choice of Migrants\*

Empirical studies in the migration literature have shown that migration enclaves (networks) negatively affect the language proficiency of migrants. These studies, however, ignore the choice of location as a function of language skills. Using data on Mexican migration to the US, we show that migrants choose smaller networks as their English language proficiency improves.

JEL Classification: F22, J15 and J61

Keywords: enclave, ethnic concentration, ethnic goods, immigrants, language skills and networks

Thomas Bauer  
IZA - Institute for the Study of Labor  
PO Box 7240  
D-53072 Bonn  
GERMANY  
Tel: (49 228) 38 94 529  
Fax: (49 228) 38 94 510  
Email: bauer@iza.org

Gil S Epstein  
Department of Economics  
Bar-Ilan University  
52900 Ramat-Gan  
ISRAEL  
Tel: (972 3) 531 8937  
Fax: (972 3) 535 3180  
Email: epsteig@mail.biu.ac.il

For further Discussion Papers by this author see:  
[www.cepr.org/pubs/new-dps/dplist.asp?authorid=127075](http://www.cepr.org/pubs/new-dps/dplist.asp?authorid=127075)

For further Discussion Papers by this author see:  
[www.cepr.org/pubs/new-dps/dplist.asp?authorid=136190](http://www.cepr.org/pubs/new-dps/dplist.asp?authorid=136190)

Ira Gang  
Department of Economics  
Rutgers University  
New Jersey Hall  
75 Hamilton St  
New Brunswick  
NJ 08901-1248  
USA  
Tel: (1 732) 932 7405  
Fax: (1 732) 932 7416  
Email: [gang@economics.rutgers.edu](mailto:gang@economics.rutgers.edu)

For further Discussion Papers by this author see:  
[www.cepr.org/pubs/new-dps/dplist.asp?authorid=116536](http://www.cepr.org/pubs/new-dps/dplist.asp?authorid=116536)

\*We thank Julie Phillips for helping us with the data and Robert Stuart for helpful discussions. This Paper was partly written while Gil Epstein and Ira Gang were visiting IZA, Bonn.

Submitted 30 July 2002

## 1. INTRODUCTION

Several papers investigate the determinants of immigrant location choice in the United States. Bartel (1989), Jaeger (2000) and Bauer, Epstein and Gang (2002), for example, find that immigrants tend to locate in cities with high concentrations of ethnically similar immigrants. Based on data from the NLSY, Borjas (1998) shows that there exist strong ethnic externalities within and across ethnic goods. His results indicate that income and human capital affect the extent of ethnic segregation in the neighborhood chosen by a household. However, the literature overlooks the effect of the immigrants' language proficiency on the location decision. Investigating the relationship between the location choice of a migrant from Mexico and his English language proficiency using data from the Mexican Migration Project, this paper aims to close this void.

Another line of literature analyzes the effect of enclaves on immigrants' educational attainment, language proficiency and labor market outcomes. Numerous empirical studies such as, among others, Borjas (1995) and Gang and Zimmermann (2000) have shown that ethnic neighborhoods have detrimental effects on the educational attainment of migrants. Chiswick and Miller (2002), using 1990 U.S. Census data on adult male immigrants from non-English speaking countries, show that linguistic concentrations (enclaves) reduce an immigrant's own English language skills. Moreover, immigrants' earnings are lower the poorer their English language proficiency and the greater the linguistic concentration of their origin language in the area in which they live. The latter result is in line with those of earlier studies such as, for example, Chiswick and Miller (1995) and Dustmann and van Soest (2001), that predominantly find a positive correlation between the language skills of migrants and their earnings. Using the 1980 U.S. Population Census, McManus (1990) shows, along the same lines, that large enclaves provide Spanish speakers better jobs for those lacking English language skills, lowering the returns to investing in English language proficiency. However, a well-recognized problem of these types of studies is the potential endogeneity of variables measuring ethnic networks in regressions of these variables on language skills, educational attainment and other labor market outcomes.

This paper looks at the location choice of migrants and connects the location decision with their language proficiency. We show that Mexican migrants to the U.S. with good English proficiency will choose, on average, a location with a small enclave whereas those with poor English proficiency choose a location with a large enclave.

## 2. DATA

We study the location choice of Mexican migrants to the United States by estimating a conditional logit model (McFadden, 1984), using data on Mexican-U.S. migration collected by the Mexican Migration Project.<sup>1</sup> The data is based on an ethno-survey approach, which combines techniques of ethnographic fieldwork and representative survey sampling. Interviews are generally conducted from December to January when sojourner U.S. migrants often return to Mexico. These interviews are supplemented with surveys of migrants located in the United States. Massey and Zeteno (1999) show that the Mexican Migration Project data represents a good source of reasonably representative retrospective data on migration to the United States.

A detailed description of the variables used in the analysis is given in Appendix A. The key variables are an indicator of the Mexican enclave in an U.S. location and three categories of the language proficiency of a migrant. Similar to other studies, the share of the total Mexican population in a particular U.S. community serves as a measure of the ethnic enclave.<sup>2</sup> The data set provides further information on the English language proficiency of a migrant for his most recent stay in the United States. For some migrants, this most recent migration experience is their first migration. We are thus able to analyze whether there are differences in network effects and language ability on the location choice of repeat versus first-time migrants. Our language proficiency variable is categorized into three levels: (*a*) does not speak nor understand English, (*b*) does not speak but understands some English and (*c*) speaks and understands English.<sup>3</sup>

In order to control for other factors that may affect location choice, the multivariate analysis includes several variables capturing the economic and social characteristics of receiving areas.<sup>4</sup> The total population in a U.S. area captures job opportunities and the general level of

---

<sup>1</sup> The Mexican Migration Project is a collaborative research project based at the University of Pennsylvania and the University of Guadalajara. See Massey, Goldring and Durand (1994), and Massey and Zeteno (1999) for descriptions of the data set. We use the MMP52 version of the data, which has information on individuals from more than 7000 households in 52 communities. In the data collection process, communities are selected based on their diversity in size, ethnic composition and economic development, not because they were known to contain U.S. migrants. Each year since 1987, two to five communities in these states are surveyed. Each community is surveyed only once. The data is made available to users at [www.pop.upenn.edu/mexmig/](http://www.pop.upenn.edu/mexmig/).

<sup>2</sup> See Phillips and Massey (2000) for the construction of this variable. We thank Julie Phillips for making this variable available to us.

<sup>3</sup> In the original data there are five levels of language proficiency. We reduce these to three categories in order to have enough observations in each language group.

<sup>4</sup> Ideally, we would like to include wages. What we would need is average wages by U.S. locations, comparable to our data set locations, for every year in our data set. This is a rather impossible task. Hence, we employ other variables (total population, unemployment rate) as proxies for wage possibilities.

economic activity in a receiving location. We also include the unemployment rate in a U.S. location to take account of both job opportunities and potential wages. Following the literature, the probability of choosing a particular location is expected to decrease with the unemployment rate in a location (see the discussion in Jaeger (2000)). Migration costs have a direct effect on location choice. In particular for migrants with low income, as is the situation for most Mexican migrants, the cost of migrating may be an important issue in determining the location choice. In order to control for these costs we include road mileage from the migrant's origin village in Mexico to the alternative U.S. locations.

Table 1 presents a description of the data we use in our analysis. Our total sample consists of 2,161 individuals from 48 Mexican villages who migrated to 47 different locations in the U.S. We assume that each person has the possibility of going to each of these specific 47 locations, but does not consider other locations. This generates 101,567 observations – each person may or may not go to each of the 47 locations. For 647 individuals the most recent U.S. migration experience was also their first trip to the U.S.; 1,514 individuals are repeat migrants. Migrants who could neither speak nor understand English dominate; they comprise 43% of all migrants, 57% of the first-time migrants but only 37% of the repeat migrants. The smallest group could not speak but understands at least some English. 32% of all migrants in our sample could speak and understand English. This fraction is relatively smaller among first-time migrants as compared to repeat migrants. Table 1 further indicates that Mexicans make up about 5.6% of the population of the U.S. locations in our sample. The network variable is slightly smaller for first-time migrants as compared to repeat migrants.

### **3. Results**

Using conditional logit analysis, we estimate several specifications. Appendix B provides a more detailed description of the econometric approach. In Table 2, Column (1) we present our basic estimation, a quadratic specification of the Mexican population share in a particular U.S. community. The estimation results show that the effect of ethnic enclaves on the probability of choosing a U.S. community follows an inverted U-shaped pattern. Up to a population share of 11%, increasing the Mexican population size increases the probability of migration to that destination. For a population share above 11% ethnic enclaves have a declining but still positive effect on the location choice; the effect becomes negative at 22%. Evaluated at the sample mean

of a Mexican population share of 5.6%, the average marginal effect of an increase in the population share by one percent is 0.47.<sup>5</sup>

Column (2) of Table 2 reports the results from a specification in which all covariates have been interacted with two dummy variables. The first dummy variable represents migrants who cannot speak but can understand English, while the second dummy variable captures migrants who can speak and understand English. The interaction terms between these language dummies and the Mexican population share and its square are statistically significant at the 1%-level, indicating that the effect of ethnic enclaves on the location choice differs among the three language groups. We also carried out log-likelihood ratio tests on the differences between the effects of the Mexican share on the probability of location choice among the three language groups. These results support the hypothesis that the effect of ethnic enclaves on the location choice is significantly different for migrants of different English language abilities. The point estimates indicate that the effect of the Mexican population share on the probability of choosing particular U.S. location peaks at 11.7% for migrants that neither speak nor understand English, at 10.6% for those that understand but do not speak English, and at 9.7% for those who speak and understand English. Hence, ethnic enclaves appear to have the strongest effect on those with the lowest language abilities and the least effect on those that have the best language proficiency.

Columns (3) and (4) of Table 2 show the results when the last model is fully interacted with dummy variables indicating whether the most current migration experience in the U.S. was the migrants' first trip and whether the migrant has been in the U.S. at least once before. The basic results presented above do not change when differentiating between first-time and repeat migrants. The effects of ethnic enclaves on the location choice for migrants with different language abilities do not differ substantially between the two types. The point estimates indicate that the probability of choosing a U.S. location peaks at an Mexican population share of 11.4% for first-time migrants who have no language abilities, at 9.8% for first-time migrants who cannot speak but understand at least some English, and at 9.3% for first-time migrants who speak and understand English. All groups are significantly different from each other. The respective numbers for repeat migrants are 11.4%, 10.7%, and 9.6%. Log-likelihood-ratio tests show that the effect of ethnic enclaves on the location choice is not significantly different for repeat migrants without any language abilities and repeat migrants that do not speak but understand at least some English. Repeat migrants with good English language abilities; however, are significantly

---

<sup>5</sup> See Appendix B for a detailed description of the calculation of the marginal effects.



different from these two groups. Note that we cannot compare between first-time and repeat migrants as there are other variables that we should control for in the repeat migration, for example, the number of repetitions.

Overall, the estimation results in Table 2 strongly indicate that there are significant differences in the location choice of migrants with differing English language abilities. Mexican migrants with poor English proficiency choose large ethnic enclaves while those with good English proficiency choose to migrate to relatively smaller enclaves.

#### **4. CONCLUSION**

Using data on Mexican-U.S. migration, this paper shows that migrants with little English proficiency choose to migrate to destinations with large ethnic enclaves. The main reason for this is that as the size of the enclave increases the need for using English decreases. Those who are proficient in English are not as dependent on network externalities provided by large enclaves. This result indicates that studies of the effect of ethnic enclaves on language abilities may suffer from endogeneity. Of course, that this is a potential issue is well known. The solution to it awaits richer data that enables the problem to be addressed.

Recently, Chiswick and Miller (2002) using the 1990 U.S. Census data on adult male immigrants from non-English speaking countries show that linguistic concentrations (enclaves) reduce an immigrant's own English language skills relative to living in smaller enclaves. Putting this result together with our results, we are able to paint a broader picture: immigrants with good English proficiency will choose to migrate to locations with relatively low concentrations of immigrants of similar ethnicity and language. As the size of the enclave is relatively small it will enable them to improve their English proficiency over time, which in turn affects their earnings and assimilation into the local population. On the other hand, immigrants with poor English proficiency will choose to migrate to locations with relatively large networks of migrants of similar ethnicity and language. This in turn decreases their ability to increase their English proficiency, which negatively affects their earnings and assimilation into the local population. We may conclude therefore that large enclaves are a potential source for a "language trap"; they attract poor proficiency English speakers and sustain their poor abilities.

## REFERENCES

- Bartel, Ann P., "Where Do the New U.S. Immigrants Live?," *Journal of Labor Economics*, 7:4 (1989), 371-391.
- Bauer, Thomas, Gil S. Epstein and Ira N. Gang, "Herd Effects or Migration Networks? The Location Choice of Mexican Immigrants in the U.S.," IZA Discussion Paper XXX (2002), <http://www.iza.org>
- Borjas, George J., "Ethnicity, Neighborhoods, and Human-Capital Externalities," *American Economic Review*, 85:3 (1995), 365-390.
- Borjas, George J., "To Ghetto or Not to Ghetto: Ethnicity and Residential Segregation," *Journal of Urban Economics*, 44 (1998), 228-253.
- Chiswick, Barry R. and Paul W. Miller, "The Endogeneity between Language and Earnings: International Analyses," *Journal of Labor Economics* 13:2 (1995), 246-288.
- Chiswick, Barry R. and Paul W. Miller, "Do Enclaves Matter in Immigrant Adjustment?," IZA Discussion Paper 449 (2002), <http://www.iza.org>
- Dustmann, Christian and Arthur van Soest, "Language Fluency and Earnings: Estimation with Misclassified Language Indicators," *Review of Economics and Statistics*, 83:4 (2001), 663-674.
- Donato, K.M., J. Durand, and D.S. Massey (1992): "Stemming the Tide? Assessing the Deterrent Effects of the Immigration Reform and Control Act," *Demography* 29(2), 139-157.
- Gang, Ira N. and Klaus F. Zimmermann, "Is Child like Parent? Educational Attainment and Ethnic Origin," *The Journal of Human Resources*, 35:3 (2000), 550-569.
- Jaeger, David A., "Local Labor Markets, Admission Categories, and Immigrant Location Choice," mimeo. (2000), <http://fsweb.wm.edu/dajaeg/research/wp/immloc.pdf>
- Massey, Douglas S., Luin Goldring, and Jorge Durand, "Continuities in Transnational Migration: An Analysis of Nineteen Mexican Communities," *American Journal of Sociology*, 99 (1994), 1492-1533.
- Massey, Douglas S., and René Zenteno, "A Validation of the Ethnosurvey: The Case of Mexico-U.S. Migration," *International Migration Review*, 34 (1999), 766-793.
- McFadden, D. (1974): "The Measurement of Urban Travel Demand," *Journal of Public Economics*, 3(4), 303-328.
- McFadden, D. (1984): "Econometric Analysis of Qualitative Choice Models," in Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics, Volume II*. Amsterdam: North-Holland.
- McManus, Walter S. (1990): "Labor Market Effects of Language Enclaves: Hispanic Men in the United States," *Journal of Human Resources*, 25(2), 228-252.
- Phillips, Julie A. and D.S. Massey, "Engines of Immigration: Stocks of Human and Social Capital in Mexico," *Social Science Quarterly* 81:1 (2000), 33-48.

**Table 1: Descriptive Statistics, Means of U.S. Recipient Locations**

		<b>Total</b>	<b>First-time Migrant</b>	<b>Repeat Migrant</b>
Unemployment Rate (in %)		7.296 (3.378)	7.256 (3.371)	7.314 (3.380)
Total Population (in 100,000)		13.617 (18.976)	13.197 (18.668)	13.796 (19.104)
Miles (in 1,000)		1,442 (0.510)	1,473 (0.518)	1,429 (0.507)
Mexican Share of Population (in %)		5.602 (6.171)	5.331 (6.113)	5.718 (6.193)
Can neither speak nor understand English	(Observations) (Individuals)	43,334 922	17,202 366	26,132 556
Cannot speak but understand English	(Observations) (Individuals)	25,803 549	6,251 133	19,552 416
Can speak and understand English	(Observations) (Individuals)	32,430 690	6,956 148	25,474 542
Total	(Observations) (Individuals)	101,567 2,161	30,409 647	71,158 1,514
Number of Mexican Villages			48	
Number of U.S. locations			47	

**Table 2: Migration Networks and Language**

	(1)	(2)	(3)	
			First-Time Migrant	Repeat Migrant
Unemployment Rate (in %)	-0.029** (0.012)	0.035** (0.016)	-0.009 (0.028)	0.057*** (0.020)
Total Population (in 100,000)	0.035*** (0.001)	0.033*** (0.001)	0.037*** (0.002)	0.030*** (0.002)
Miles (in 1,000)	-0.150** (0.065)	-0.149 (0.096)	-0.443*** (0.152)	0.041 (0.125)
Mexican Share of Population (in %)	0.463*** (0.017)	0.396*** (0.023)	0.364*** (0.037)	0.412*** (0.029)
Mexican Share of Population (in %) <sup>2</sup>	-0.021*** (0.001)	-0.017*** (0.001)	-0.016*** (0.002)	-0.018*** (0.002)
<i>Interactions with cannot speak but understand:</i>				
Unemployment Rate (in %)	-	-0.116*** (0.030)	-0.076 (0.064)	-0.141*** (0.034)
Total Population (in 100,000)	-	0.002 (0.002)	0.012*** (0.004)	0.000 (0.002)
Miles (in 1,000)	-	0.095 (0.162)	-0.173 (0.336)	0.051 (0.193)
Mexican Share of Population (in %)	-	0.135*** (0.046)	0.243** (0.100)	0.079 (0.052)
Mexican Share of Population (in %) <sup>2</sup>	-	-0.008*** (0.003)	-0.015** (0.006)	-0.005 (0.003)
<i>Interactions with can speak and understand:</i>				
Unemployment Rate (in %)	-	-0.144*** (0.029)	-0.185*** (0.069)	-0.152*** (0.033)
Total Population (in 100,000)	-	0.007*** (0.002)	0.010** (0.004)	0.008*** (0.002)
Miles (in 1,000)	-	-0.005 (0.154)	0.329 (0.326)	-0.207 (0.183)
Mexican Share of Population (in %)	-	0.203*** (0.046)	0.250** (0.100)	0.182*** (0.053)
Mexican Share of Population (in %) <sup>2</sup>	-	-0.014*** (0.003)	-0.017** (0.007)	-0.013*** (0.003)
Log-Likelihood	-6204.01	-6154.57	-6124.76	
Pseudo-R <sup>2</sup>	0.254	0.260	0.264	

Notes: Observations: 101,567; individuals: 21,161. Standard errors in parentheses. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

## Appendix A: Data description

### *Total Population:*

Data for selected years between 1970 and 1995 were obtained from periodic Census publications, such as the CPS and County and City Yearbook. Data were obtained for the following years: 1970, 1974, 1976, 1977, 1980, 1984, 1986, 1987, 1990, and 1991. The population for the intercensal years was estimated by assuming an exponential growth function. To estimate the population between 1992-1995, the constant growth rate that prevailed between 1980 and 1991 was applied. Source: *Mexican Migration Project 52*.

### *Unemployment Rate:*

The most recent information on the number unemployed and the size of the civilian labor force at the county level was obtained for the years 1974 and 1976-1996 from the Bureau of Labor Statistics, Local Area Unemployment Statistics Division. For the early 1970s, no information by county is available although information on unemployment for the censal years 1960 and 1970 is available. For the years 1971-1973, the assumption was made that unemployment rates in a county follow the same trends as that of the state. An estimate of the unemployment rate for 1975 was obtained by averaging the unemployment rates for 1974 and 1976. Source: *Mexican Migration Project 52*.

### *Migration Costs:*

We collected data on three measures of migration costs. For *Miles* and *Hours* we entered in the main town in the Mexican state in which the origin village is located and the main town in the U.S. location into *Mapquest* ([www.mapquest.com](http://www.mapquest.com)) and into *Mapblast* ([www.mapblast.com](http://www.mapblast.com)). For *Actual Costs* the data come from the *Mexican Migration Project 52*. Since the actual cost data was very sketchy, we decided not to use it. Trials with the *Hours* and the *Actual Costs* data yielded similar results to those when we used *Miles*.

### *Mexican Share of Population:*

This variable has been obtained from the U.S. Census Bureau for the censal years 1970, 1980 and 1990. A second-degree polynomial equation was estimated to these three data points to estimate the size of the Mexican foreign-population in each area during the inter-censal years. To estimate the Mexican foreign-born population in the years 1991-1995, it has been assumed that the annual growth rate during this period is the same as the annualized constant growth rate in each area between 1980 and 1990. The size of the Mexican foreign-born population is then divided by the *Total Population* in a U.S. location. Source: We are very grateful to Julie A. Phillips for making this variable available to us.

### *Number of Current Migrants:*

To calculate these variables we make use of an event-history file provided by the Mexican Migration Project. This event-history file contains detailed labor and family histories of each household head, for each year from the birth of the household head until the year of the survey. Donato, Durand and Massey (1992) give a description of the event-history file. Using this event history file, we calculated for each year  $t$  the number of current migrants from the Mexican community  $k$  in the U.S. community  $j$ .

### *English Language Ability:*

For each household head the Mexican Migration Project includes self-reported information on his language ability on his last trip to the U.S. This variable provides information on the language ability in five categories: (1) Doesn't Speak nor Understand English, (2) Doesn't Speak but Understands Some English, (3) Doesn't Speak but Understands Well, (4) Speaks and Understands Some English, (5) Speaks and Understands English Well. For our estimations we combined categories (2) and (3) and categories (4) and (5) in order to have enough individuals in each language group.

## Appendix B: Econometric Model

In our econometric analysis we estimate a conditional logit model (McFadden, 1974, 1984). Each Mexican migrant  $i$  faces a choice among  $J$  alternative U.S. communities. Assume that the utility of choosing location  $j$  is given by

$$U_{ij} = Z_j \gamma + \varepsilon_{ij}, \quad (1)$$

where  $Z_j$  is a vector of the characteristics of the U.S. community  $j$  which includes the network variables, and  $\varepsilon_{ij}$  is an error term that is assumed to be independent and identically distributed with a Weibull distribution.  $\gamma$  is a parameter vector to be estimated. Individual  $i$  is assumed to be utility maximizing. The probability that an individual  $i$  choose community  $j$  is given by

$$\Pr(U_{ij} > U_{ik}) \quad \text{for all } k \neq j. \quad (2)$$

Let  $Y_i$  be a random variable that takes the values 0 and 1 indicating the location choice made by the migrant. The probability that individual  $i$  choose the U.S. community  $j$  can then be written as

$$\Pr(Y_i = j) = \frac{\exp(Z_j \gamma)}{\sum_{j=1}^J \exp(Z_j \gamma)}. \quad (3)$$

Equation (3) can be estimated using maximum likelihood. Note that our sample is restricted to individuals who actually migrated at some point in time to the U.S. The analysis does not consider migration within Mexico.

Since individual characteristics of a migrant, such as language abilities, are the same for all choices, they fall out of the probability described in equation (3). Hence, only independent variables describing the attributes of the U.S. communities can enter the model. We are, however, interested whether migration networks affect the location choice of individuals with different language abilities differently. Therefore, we estimated an interacted model, i.e.:

$$\Pr(Y_i = j) = \frac{\exp(Z_j \gamma_1 + L_{i1} Z_j \gamma_2 + L_{i2} Z_j \gamma_3)}{\sum_{j=1}^J \exp(Z_j \gamma_1 + L_{i1} Z_j \gamma_2 + L_{i2} Z_j \gamma_3)}, \quad (4),$$

where  $L_{i1}$  is a dummy variable that takes the value 1 if an individual  $i$  does not speak but understands some English, and  $L_{i2}$  is a dummy variable that takes the value 1 if an individual  $i$  can speak and understands some English and 0 otherwise.

Comparing the estimated coefficients  $\gamma_1$ ,  $\gamma_2$  and  $\gamma_3$  allows us to test whether the effects of networks on the location choice of migrants differ by the language ability of the migrants. Based on the available data these variables have been constructed for every household head for the last time they have been in the U.S. (see Appendix A for a detailed description). It could be expected, however, that the relationship between language, migration networks and location choice differ between individuals for whom this last trip was the first trip to the U.S. and migrants that have already been in the U.S. before. To test this hypothesis, we fully interacted equation (4) with two additional dummy variables, one variable indicating whether it was a persons' first trip and one variable indicating whether a person has been in the U.S. already before, i.e. we estimate

$$\Pr(Y_i = j) = \frac{\exp(T_{in} Z_j \gamma_1 + T_{in} L_{i1} Z_j \gamma_2 + T_{in} L_{i2} Z_j \gamma_3)}{\sum_{j=1}^J \exp(T_{in} Z_j \gamma_1 + T_{in} L_{i1} Z_j \gamma_2 + T_{in} L_{i2} Z_j \gamma_3)}, \quad n=1,2,$$

where  $T_{i1}$  is a dummy variable that equals 1 if the most current trip to the U.S. has also been a migrants' first trip, and  $T_{i2}$  a dummy variable that equals 1 if the individual has been in the U.S. already before.

The marginal effects of a change in the characteristics  $Z_j$  of a U.S. location  $j$  on the probability that a Mexican migrant will choose location  $j$  are given by the derivative of equation (3) with respect to the characteristics  $Z_j$ . Note that these marginal effects will vary with the characteristics of a U.S. location  $j$ . Therefore, we follow the approach chosen by Jaeger (2000) and calculate average effects of a change in the characteristics  $Z_j$  on  $\Pr(Y_i = j)$ , i.e.  $\partial \Pr(Y_i = j) / \partial Z_j = [(1/J)(1 - (1/J))] \hat{\beta}$ , where  $J=47$ . Hence, to obtain average marginal effects, the coefficients reported in Table 2 have to be multiplied by 0.0208.